A Biomedical Image Retrieval Framework Based on Classification-Driven Image Filtering and Similarity Fusion

Md Mahmudur Rahman, Sameer K. Antani, George R. Thoma
U.S. National Library of Medicine,
National Institutes of Health,
Bethesda, MD, USA

Abstract

This paper presents a classification-driven biomedical image retrieval approach based on multi-class support vector machine (SVM) and uses image filtering and similarity fusion. In this framework, the probabilistic outputs of the SVM are exploited to reduce the search space for similarity matching. In addition, the predicted category of the query image is used for linear combination of similarity. The method is evaluated on a diverse collection of 5000 biomedical images of different modalities, body parts, and orientations and shows a halving in computation time (efficiency) and 10% to 15% improvement in precision at each recall level (effectiveness).

1 Introduction

Quality and speed of content-based image retrieval (CBIR) from large biomedical image collections can be improved by reducing the search space by filtering out irrelevant images and learning about the image categories. For example, to search posteroanterior (PA) chest x-rays with enlarged heart database images can be pre-filtered with automatic categorization according to modality (e.g., x-ray), body part (e.g., chest), and orientation (e.g., PA). Next, similarity matching can be performed between query and target images in the filtered set to find "enlarged heart" as a distinct visual property.

Prior work has explored medical image classification into multiple semantic categories for effective retrieval [1, 2]. For example, automatic categorization of 6231 radiological images into 81 categories is examined in [2] by utilizing a combination of global texture features applied to down-scaled images and K-nearest-neighbors (KNN) classifier. Although these approaches demonstrate promising results for medical image classification at a global level, they do not directly relate classification to retrieval, instead only suggest its potential for application to image annotation and pre-filtering.

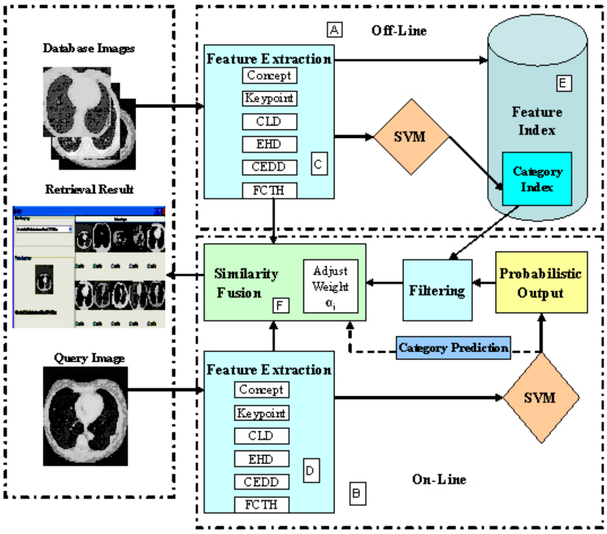

Figure 1. Image Retrieval Framework

To minimize limitation of the low-level feature representations resulting in poor retrieval quality due to the semantic gap and motivated by the successful use of machine learning for CBIR, we present a learning-based retrieval framework that uses a novel image filtering and similarity matching method. In this framework, several image features at different levels of abstraction are extracted for classification and retrieval. First the image collection is grouped into various categories using probabilistic outputs of a multi-class Support Vector Machine (SVM) [3]. Next on-line category prediction of the query image is used to filter out irrelevant class images, and then pre-computed category-specific feature weights are applied in a linearly combined similarity-matching function.

The block diagram of the proposed image retrieval framework is shown in Fig. 1 with different modules. The rest of the paper is organized as follows. Section 2 describe the image categorization approach at a global level by utilizing the multi-class SVM. Section 3 presents the image filtering approach based on the probabilistic outputs of the classifiers. The similarity fusion approach based on image classification is presented in Section 4. The experiments and the analysis of the results are presented in Sections 5 and 6, respectively.

2 Multi-Class SVM-based Categorization

In this research, we utilize a multi-class classification method by combining all pairwise comparisons of binary SVM classifiers, known as *one-against-one* or pairwise coupling (PWC) [4]. PWC constructs binary SVM's between all possible pairs of classes. Hence, for M classes, this method uses $M * (M-1)/2$ binary classifiers that individually compute a partial decision for classifying a data point (image). During the testing of a feature $\mathbf{x}$, each of the $M * (M-1)/2$ classifier votes for one class. The winning class is the one with the largest number of accumulated votes.

The input feature vector set of training images is manually annotated with a single global category label ω_i out of a set of M labels: $\{\omega_1, \cdots, \omega_i, \cdots, \omega_M\}$. In this context, given a feature vector $\mathbf{x}$, the multi-class SVM estimates the probability or confidence scores of each category as

$$p_m = P(y = \omega_m \mid \mathbf{x}), \text{for } 1 \leq \text{m} \leq \text{M} \quad (1)$$

The final category of a feature is determined based on the maximum probability score.

3 Image Filtering

We use image filtering to reduce the search space based on the output of the classification approach above as follows

$$\mathbf{p}_j = [p_{j_1}, \cdots, p_{j_m}, \cdots, p_{j_M}]^{\mathrm{T}} \quad (2)$$

Here, $p_{j_m}, 1 \leq m \leq M$, denotes the probability or class confidence score that an image I_j belongs to the category ω_m.

During the off-line indexing process, this output is stored as the category vector of the database images in a *category index* along with the feature indices. Similar feature extraction and category prediction stages are performed on-line when the system is searched using an unknownquery image. The category vector of a query image I_q (Equation (2) and the vectors of the database images from the category index are evaluated to identify candidate target images in the collection, thereby filtering out irrelevant images from further consideration. To minimize misclassification errors, instead of only considering the image categories based on the highest obtained probability values, $n < M$ nearest classes of the target images to the query image are considered.

The process validates for class overlap between the query and target images. Generally, the value of $n << M$ to prevent inclusion of distant classes and provide effective filtering. A target image is only selected for further matching if at least one common category is found out between the top n categories of the query image and itself. This further reduces the risk of searching wrong images due to misclassification. Steps of the filtering algorithm are presented below.

Algorithm 1 Classification-Based Image Filtering

(Off-line): Select a set training images of M categories with associated category label for SVM learning.
(Off-line): Store the category vectors (Equation 2) of N database images as a category index.
(On-line): For a query image I_q, determine the category vector as $\mathbf{p}_q = [p_{q_1}, p_{q_2}, \cdots, p_{q_M}]^{\mathrm{T}}$.
for $j = 1$ to N **do**
  Consider the top ranked $(n < M)$ category labels for I_q and I_j after sorting the elements in the category vectors.
  Construct the category label sets as S_q and S_j for the top ranked categories of I_q and I_j respectively. Here, $|S_q| = n$ and $|S_j| = n$.
  if $(S_q \cap S_j \neq \emptyset)$ **then**
    Consider I_j for further similarity matching.
  end if
end for

4 Category-Specific Similarity Fusion

Typically in CBIR the most commonly used similarity matching method is a linear combination of different low-level image features with pre-determined weights. In this approach, for example, a texture feature will have the same weight as color for the search of the microscopic pathology, x-ray, ultrasound, CT, or MRI images, even though they may have a negligible presence in some of them. This can often result in poor retrieval quality.

In this framework, the similarity between a query image I_q and target image I_j is described as

$$\text{Sim}(I_q, I_j) = \sum_{F} \alpha^F \, \text{S}^F(I_q, I_j) \quad (3)$$

where F represents individual and $\text{S}^F(I_q, I_j)$ is the similarity matching function and α^F are weights for the different image features within our framework. We use a fusion-based linear combination scheme with pre-computed

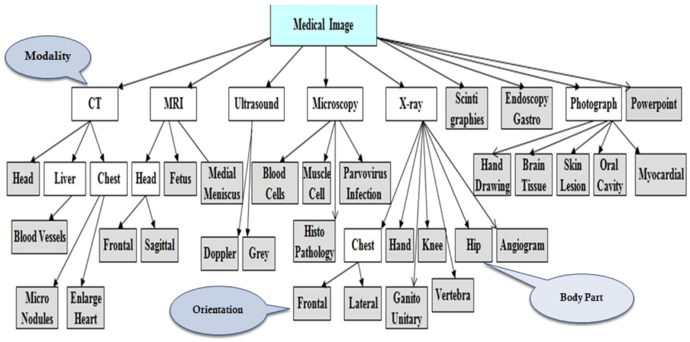

Figure 2. Global image classification structure

category-specific feature weights based on the on-line category prediction of a query image.

5 Evaluation

We evaluate our method on a biomedical image collection that comprises 5000 images in 30 manually-assigned disjoint global categories. The collection is a subset of a larger collection of 77,000 images made available by the medical image retrieval track in 2010 [9] of ImageCLEF [1] evaluation. The images are further classified into three levels (e.g., imaging modalities, body parts, and orientations or distinct visual observation), as shown in Figure 2.

In this work, images are presented with "local concepts" that comprise of color and texture patches from local image regions. The example of local patches are, homogeneous texture patterns in radiological images, differential color and texture structures in microscopic pathology and dermoscopic images, etc. The variation in the local patches is modeled as local concepts by using SVM learning and used to represent the images as a vector of concepts [5]. We also extracted Lowe's [6] Scale-Invariant Feature Transform (SIFT)-based local descriptor that transforms the image information related to the local features in a set of scale-invariant coordinates. The calculated features are vector quantized (features of interest points are converted into visual words or keypoints) and images are finally represented by a bag of quantized features (e.g., bag of keypoints). In addition, the Color Layout Descriptor (CLD) and Edge Histogram Descriptor (EHD) of MPEG-7 standard [7] and Color Edge Direction Descriptor (CEDD) and Fuzzy Color Texture Histogram (FCTH) from the Lucene image retrieval (LIRE) library [8] are used to represent images.

A training set of about 2400 images is used for SVM learning to categorize images at a global level. The images

[1] http://www.imageclef.org/

Table 1. Individual classifier accuracy

Concept	SIFT	EHD	CLD
71%	63%	52%	52%
FCTH	**CEDD**	**Combined**	
63%	69%	79%	

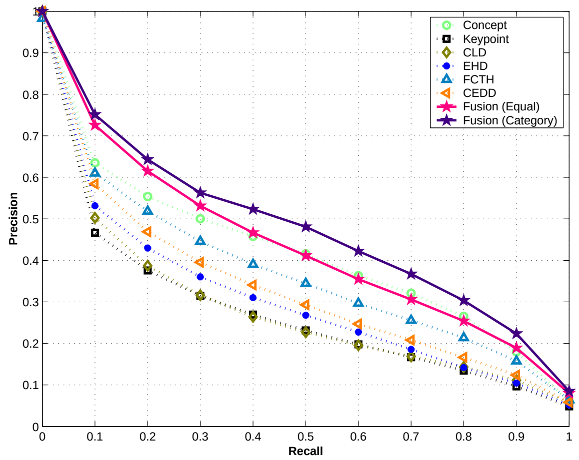

Figure 3. Precision-Recall graph

are classified into one of 8 modalities, viz., Graphic, nuclear medicine, CT, MR, XR, PET, ultrasound, x-ray, and optical images (photograph)with approximately 300 images per class. There is considerable intra-class heterogeneity in this classification. For instance, the optical image class contains both microscopic images as well as photographs, and PET image class contains both PET and PET/CT.

6 Results

To measure the classification accuracies, we utilize a test set of 2620 images. From Table 1, we observe that the best classification accuracy rate of 79% is achieved in the combined feature space. For a quantitative evaluation of the retrieval results, we selected all the images in the collection as query images and used *query-by-example (QBE)* as the search method. A retrieved image is considered a match if it belongs to the same category as the query image out of the 32 disjoint categories. Fig. 3 shows the precision-recall (PR) curves of the individual feature spaces when the similarity matching is performed using the Euclidean distance measure. Their performance is compared to a similarity fusion approach by providing equal weight to each feature (e.g., Fusion (Equal)) and adjustable weights of each feature based on category prediction (e.g., Fusion (Category)) in a linear combination. By analyzing Fig. 3, we see that the best performance is obtained in terms of precision at each recall level when the similarity scores are combined

Table 2. Average retrieval time

Search Method	Filtering	Time (ms)
Fusion (Equal Weight)	No	600
Fusion (Equal Weight)	Yes	360
Fusion (Category)	No	630
Fusion (Category)	Yes	420

for individual features based on category-specific weights. In general, we achieved around 10at most recall levels (0.1-0.9) when our proposed similarity fusion approach is compared to the equal weight-based similarity fusion.

Finally, to test the effectiveness and efficiency of the prefiltering approach discussed in Section 1, retrieval is performed with and without filtering on different fusion-based approaches. For filtering the 3 top ranked (e.g., n = 3) category labels are considered for both query and target images due to its better performance. By applying the similarity fusion approaches in the filtered image set, we achieved exactly the same precision when compared to the linear search scheme as shown in Fig. 3. There is no decrease in retrieval accuracy due to the fact that the filtering algorithm only discards those images from further consideration that are perceptually dissimilar to the query images to a significant extent.

On the other hand, the major gain in searching on a filtered image set is that it takes less computational time compared to a linear search on the entire collection. Hence, to test the efficiency of the fusion-based search schemes, we compared the average retrieval time (in milliseconds) of query images with and without applying the filtering scheme. The experiment is performed in an Intel Pentium Dual-Core CPU clocked at 3.40 GHz and 4 GB of RAM with Microsoft Windows XP Professional operating system. Table 2, shows that the linear search time is almost double for different fusion schemes when compared to the search in the filtered image set. Hence, the proposed image retrieval framework with filtering and fusion has proved to be both effective and efficient.

7 Conclusion

In this paper, a novel learning-based and classification-driven image retrieval framework is proposed for biomedical image retrieval from diverse medical image collections. Unlike other approaches we consider effectiveness of image type classification to retrieval. The retrieval performance results are promising and clearly show the advantage of similarity fusion and image modality filtering. We are also investigating other machine learning methodologies to complement SVM.

Acknowledgments

This research is supported by the Intramural Research Program of the National Institutes of Health (NIH), National Library of Medicine (NLM), and Lister Hill National Center for Biomedical Communications (LHNCBC). We thank the ImageCLEFmed [9] organizers for making the database available for research .

References

[1] A. Mojsilovis, J. Gomes, "Semantic based image categorization, browsing and retrieval in medical image databases," *Proc. IEEE Internat. Conf. on Image Processing*, vol. 3, pp. 145–148, 2002.

[2] T.M. Lehmann, M.O. Güld, T. Deselaers, D. Keysers, H. Schubert, K. Spitzer, H. Ney, and B.B. Wein, "Automatic categorization of medical images for content-based retrieval and data mining," *Computerized Medical Imaging and Graphics*, vol. 29, pp. 143–155, 2005.

[3] V. Vapnik, *Statistical Learning Theory*. New York, NY, Wiley; 1998.

[4] T.F. Wu, C.J. Lin, R.C. Weng, "Probability Estimates for Multi-class Classification by Pairwise Coupling.", *J. of Machine Learning Research*, vol. 5, pp. 975–1005, 2004.

[5] M.M. Rahman, S.K. Antani, and G.R. Thoma, "A Medical Image Retrieval Framework in Correlation Enhanced Visual Concept Feature Space," *22nd IEEE International Symposium on Computer-Based Medical Systems (CBMS)*, August 3-4, 2009, Albuquerque, New Mexico, USA.

[6] D.G. Lowe, "Distinctive image features from scale-invariant keypoints", *International Journal of Computer Vision*, vol. 60(2), pp. 91–110, 2004.

[7] S.F. Chang, T. Sikora, and A. Puri, "Overview of the MPEG-7 standard," *IEEE Trans Circ Syst Video Technology*, vol. 11, pp. 688–695, 2001.

[8] M. Grubinger, P. Clough, A. Hanbury, and H. Müller, "Lire: lucene image retrieval: an extensible java CBIR library," *Proceeding of the 16th ACM international conference on Multimedia*, Vancouver, British Columbia, Canada, pp. 1085-1088, 2008.

[9] H. Müller, J. Kalpathy-Cramer, I. Eggel, S. Bedrick, C.E. Kahn, Jr., and W. Hersh, "Overview of the CLEF 2010 medical image retrieval track," *Working Notes for the CLEF 2010 Workshop*, Padova, Italy, 2010.